REGIONAL SOCIAL ACCOUNTING MATRICES FOR THE EU27 (IOTNUTS2)

MARC MÜLLER
Center for Development Research (ZEF)
Walter-Flex-Straße 3 Bonn, D-53113-Germany
Email: marc.mueller@ilr.uni-bonn.de

EMANUELE FERRARIb
European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS)
Calle Inca Garcilaso, 3 -Seville, 41092 – Spain
Email: emanuele.ferrari@ec.europa.eu

Selected Poster prepared for presentation at the International Association of Agricultural Economists (IAAE) Triennial Conference, Foz do Iguaçu, Brazil, 18-24 August, 2012.

REGIONAL SOCIAL ACCOUNTING MATRICES FOR THE EU27 (IOTNUTS2)

MARC MÜLLER[a], EMANUELE FERRARI[b]

[a] *Center for Development Research (ZEF)* [b] *European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS)*

WORK IN PROGRESS: PRELIMINARY DRAFT.
PLEASE DO NOT QUOTE WITHOUT PRIOR AGREEMENT WITH THE AUTHORS.

Abstract:
Agricultural polices in the EU are increasingly targeting not only the agricultural sector but also other economic branches. The indirect effects of these policies, as the rural development ones, might be as important as the direct ones, mainly on factor markets as labour. In addition, in order to better scale the adopted agricultural measures, policy makers are devoting more attention to the regionalized impacts of these policies. For these reasons, a pure partial-equilibrium agricultural model is not enough to account the effects of the EU agricultural policies. The development of regionalized Computable General Equilibrium models and the linkages with already developed regionalized agricultural partial equilibrium models is a fundamental step for agricultural economists. The greatest challenge to build a regional general equilibrium model for all EU27 NUTS2 regions is the database construction. This work show the main steps needed to construct such a database, called IOTNUTS2.

Keywords: regional input-output tables, non-survey method, location quotient, European Union

JEL classifications: D57, R10, O52

1. Introduction[1]

The reforms of the Common Agricultural Policy (CAP) is creating increasing interest amongst policy makers and analysts in the backward- and forward linkages between regional agricultural and non-agricultural sectors, related labour markets, and regional development policies. Thus, demand for model-based analyses of regional development policies in a multi-sector context is increasing in order to capture the effects of these policies on all braches of the economy and to allow a better regional scaling of agricultural policies. Mattas et al. (2011), in their analysis for 5 regions in the EU, highlight that an inspection of regional Input-Output tables (IOTs) reveals deep differences in the structure of the regional economy. The size of same sectors and the distribution of multipliers diverge among the regions. According to Mattas et al. (2011) this is a clear indication that that Pillar II programmes to be effective and boost the regional economy should be highly flexible. If this is true for a sample of 5 regions, the differences are magnified for the all set of regions.
Addressing regional heterogeneity requires multi-sector data on a sub-national scale. Such datasets are usually not sufficiently detailed, if available, which gave rise to numerous non-survey methods to generate regional IOTs based on combinations of regional indicators and national datasets. Many examples of regionalization of national tables for single or multiple

[1] The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

regions (QUOTE SOME???) exist in the literature. In addition, at national level some tentative to construct consistent regionalized tables have been pursued mainly by National Statistical Offices (NSO) following survey-based method (i.e. Finland, OFS) or national research institutes following non survey-based methods (Fritz et al., 2003) and link them to multi-sectoral regionalized national models (Casini-Benvenuti and Paniccíá, 2003). At the best of our knowledge, a complete set of SAMs for all the EU NUTS2 does not yet exist and this work fulfil this deficiency in the literature.
The goal of this paper is to describe the steps to build a database of Social Accounting Matices (SAM) on NUTS2[2] level, called IOTNUTS2. Firstly we illustrate the inventory on regional datasets for EU27: national and regional databases from Eurostat (ESTAT) and the significant informational gain attained with the datasets coming from Member States NSOs. Following standard non-survey procedures, the data are then combined to populate the regional SAMs. Survey-based regional tables coming from NSOs are used to test the reliability of the techniques adopted in this work to combine national and regional datasets. This test shows that for the majority of economic sectors, non-survey methods generate reliable substitutes for otherwise collected indicators. These matrices are then balanced following a modified Stone-Byron method.

SECONDLY….FINAL POINT ON CLASSIFICATION

The importance of this database is two-folded. A SAMs based dataset permits to exploit the characteristics of Computable General Equilibrium analyses. Firstly, this allows to asses the feedbacks of agricultural policies on non-agricultural sectors and on the factor markets. Policies like reforestation programmes, the promotion of investment in agro-tourism or environmental services, and the support for the production of renewable energy by farming enterprises and all the policies related to the so-called Pillar II of the CAP can be regionally modelled. Such measures primarily target the agricultural sector, but are likely to influence other economic sectors and aggregate regional income, depending on the regional economic structure and the dominance of agriculture. Secondly, the construction of such a regional database allows a linkages between modified existing regionalized national Computable General Equilibrium (CGE) models (i.e. RegFin/RegPol - Törmä and Zawalinska, 2007) and partial equilibrium models like the "Common Agricultural Policy Regionalized Impact analysis modelling system" (CAPRI – Britz and Witzke, 2008), which already covers all the EU NUTS2 regions.
The advantages of both types of modelling, the generality of the CGE and its capacity of taking into account all the aspects of an economy and the "depth" of the PE model and the abundance of details in the modelling of a single sector, can be exploited in three main ways (Törmä et al. 2010). They might be integrated through their database (Mueller et al., 2009) or with a sequential implementation of scenarios, where one the results of a model serve as input for other models (Nowicki et al., 2009 for the Scenar2020 Project). A third approach, which is allowed only with the development of a database like the IOTNUTS2, is the iterative calibration of structural model parameters, as developed within the SEAMLESS project

[2] The Nomenclature of Territorial Units for Statistics (NUTS) classification is a hierarchical system for dividing the economic territory of the EU for the purpose of collection, development and harmonisation of EU regional statistics and socio-economic analyses of the regions. The regional classification follows this hierarchy: NUTS 1 (major socio-economic regions), NUTS 2 (basic regions for the application of regional policies), and NUTS 3 (small regions for specific diagnoses). The current NUTS classification lists 97 regions at NUTS 1, 271 regions at NUTS 2 and 1303 regions at NUTS 3 level.
(http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction)

(Jansson et al. 2009). This approach ensured the harmonized simulation behaviour of the models for matching endogenous variables.
The rest of the paper is organized as follows. In Section 2 the SAMs structure is illustrated. Sections 3 and 4 illustrate the compilation of the starting database with information coming from ESTAT and National Statistical Offices. Section 5 evaluates the performances of the non-survey methods adopted to compile the regional tables. In section 6 the full set of SAMs is derived and balanced while section 7 concludes.

2. Structure of the Social Accounting Matrices and Satellite Accounts

The development of a European-wide database of SAMs on regional scale according to the NUTS nomenclature is a prerequisite for the usage of regionalised CGE models for quantitative evaluation of the EU regional policies. The major challenges to build such a database are the lack of regional data and the high sectoral aggregation in regional branch accounts provided by ESTAT.
The targeted database consists of numerous sub-tables, classified into core-SAM accounts and satellite or auxiliary accounts. The core datasets appear in the final SAMs, such as intermediate demand distinguished by origin and economic branch or compensation of employees by branch. Satellite accounts serve as control-totals for the core datasets, i.e. gross value-added does not appear in the final SAM, but represents “compensation of employees” plus “net taxes on production” plus “operating surplus”.
A crucial feature of the database is the distinction of some items by their origin, i.e. if they are produced in the same region in which they are consumed, or imported from a different region in the same country, or imported from abroad. The core SAMs and sub-matrices are summarized in Appendix 1.A set of satellite accounts contain the type of information that is not directly part of the SAMs, but control for the required SAM entries. Satellite accounts and the correspondences with the SAM are illustrated in the Appendix 2.
To ensure the database consistency with the ESA95 standards and to facilitate communication with the Member States’ statistical departments, the branch classification for the target database follows the Nomenclature for Economic Activities (NACE) within the European System of National Accounts (ESA95) at first digit (16 branches), with the exceptions of Agriculture, Hunting and Forestry (NACE: A) and Manufacturing branches (NACE: D). These two sectors are further disaggregated into two (for agriculture) and three (for manufacture) sub-branches (Table 1). Therefore, matrices contain 19 branches as elements, but data are also collected for the aggregates A and D. Primary factors (f) contains the elements “Labour”, “Land”, and “Physical capital”. Trade partners (w) for each region are either the intra-national markets for trade between regions or the external markets for international trade between countries. Regional consumers are regional governments and households.

Table 1 Target Branches of the IOTNUTS2 Database

Block	Code	Description
Target sectors	AA01	"Agriculture, hunting and related services"
	AA02	"Forestry, logging and related services"
	B000	"Fishing"
	C000	"Mining and quarrying"
	DA00	"Food products, beverages, and tobacco"
	DF00	"Coke, refined petroleum products, and nuclear fuels"
	DZ00	"Other manufacturing"
	E000	"Electricity, gas and water supply"

Block	Code	Description
	F000	"Construction"
	G000	"Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods"
	H000	"Hotels and restaurants"
	I000	"Transport, storage and communication"
	J000	"Financial intermediation"
	K000	"Real estate, renting and business activities"
	L000	"Public administration and defence; compulsory social security"
	M000	"Education"
	N000	"Health and social work"
	O000	"Other community, social, personal service activities"
	P000	"Activities of households"
Sector aggregates (NACE16 and NACE06)	A000	"Agriculture, hunting and forestry"
	D000	"Manufacturing"
	A2B	"Agriculture, hunting, forestry and fishing"
	C2E	"Total industry (excluding construction)"
	F00	"Construction"
	G2I	"Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication"
	J2K	"Financial intermediation; real estate, renting and business activities"
	L2P	"Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies"

3. Regional and National Data from ESTAT

The first step in the compilation of this inventory is the review of national and regional datasets from ESTAT.

For the year 2005, 21 Member States provide national IOTs in the desired format (i.e. National Account Input Output, NAIO Tables 1800 and 1900). As a full set of IOTs is available for the year 2000 (either from ESTAT or Rueda-Cantuche et al. 2009), the first step is to update to 2005 the IOTs for the missing Member States. This can be done by using a completed set of annual national accounts (NAMA) for the relevant branch indicators. These datasets provide time series for indicators as “b1g: gross value-added at basic prices” or “p1: output at basic prices” and can be used to adjust the 2000 IOTs for the targeted base-year 2005.

National income is distributed in the national SAMs across the domestic institutions and the “Rest of the World” by combining updated IOTs and the annual sector accounts (NASA) dataset. NASA contains data on flows between sectors, domestic institutions, and the 'rest or the world' and is the only source of information for factor incomes from abroad, transfers received by households and direct taxes paid by enterprises and households. Critical Members States are Luxembourg, Cyprus, and Malta, as no NAIO, NAMA or NASA datasets are fully available. The problem is limited as these three Member States consist of only one NUTS2 region.

The “Regional statistics (reg)” section of ESTAT covers a wide range of indicators. Attention is devoted to those sub-sections which provide information for the structure of economic branches in the NUTS2 regions. The regional branch accounts have, despite their rough representation of economic branches at NACE06[3], full coverage for all 271 NUTS2 regions in

[3] A_B: "Agriculture, hunting, forestry and fishing", C_E: "Total industry (excluding construction)", F: "Construction", G_H_I: "Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication", J_K: "Financial

2005 for the indicators “value-added at basic prices (b1g)” and “total employment (emp)”. This dataset represents therefore the most valuable asset for the subsequent compilation steps. A breakdown to 19 branches could be achieved by using the NAMA and NAIO datasets on a national scale for the indicators "compensation of employees (d1)", b1g, and emp. Structural Business Statistics and Regional Economic Accounts for Agriculture (EAA) data are then used as supplements. The structural business indicators cover the NACE Rev 1.1 sections C to K, with a breakdown of branches at the 2-digit. This dataset provides figures on employment, wages and salaries.
As agriculture is not covered neither in the branch accounts nor in the structural business statistics EAA is also evaluated. Full coverage for “compensation of employees” and “gross value-added” is given for 19 Member States in 2005, while coverage above zero and below 100% can be observed for the Czech Republic, and Italy has full coverage for b1g, but zero coverage for d1. The remaining countries have no entries.

4. National Statistical Organisations

After screening the datasets from ESTAT, we browse the homepages and contact all the national statistical institutes of the Member States. In general, we discover that the national classification schemes follow closely – but not fully – the ESA95 system.
To evaluate and compare the gain of information from contacting the Member States' statistical institutions, the additional data points across all branches are summarized in an indicator. Under the assumption that the regional datasets are consistent with the ESTAT branch accounts, one would need (na6-1) additional data points to construct the full branch accounts (e.g. if data on a01 and a02 are available, then the remaining entry for b could be obtained residually, provided that a01+a02 < A2B).
Then we construct an “informational gain” indicator over all branches (TIG) for all regional branches AR. G^{19_AR} stands for the aggregator from regional to IOTNUTS2 branches for the respective datasets I and G^{19_6} represents the aggregator matrix from 19 to 6 branches

$$TIG_i^{MS,R} = \sum_{A6}\left(\sum_{b19}\left[G_{b19,A6}^{19\_6}\cdot\delta\left(\sum_{AR}G_{i,b19,AR}^{19\_AR}\right)\right]-1\right)\Big/\sum_{A6}\left(\sum_{b19}\left[G_{b19,A6}^{19\_6}\right]-1\right) \tag{1}$$

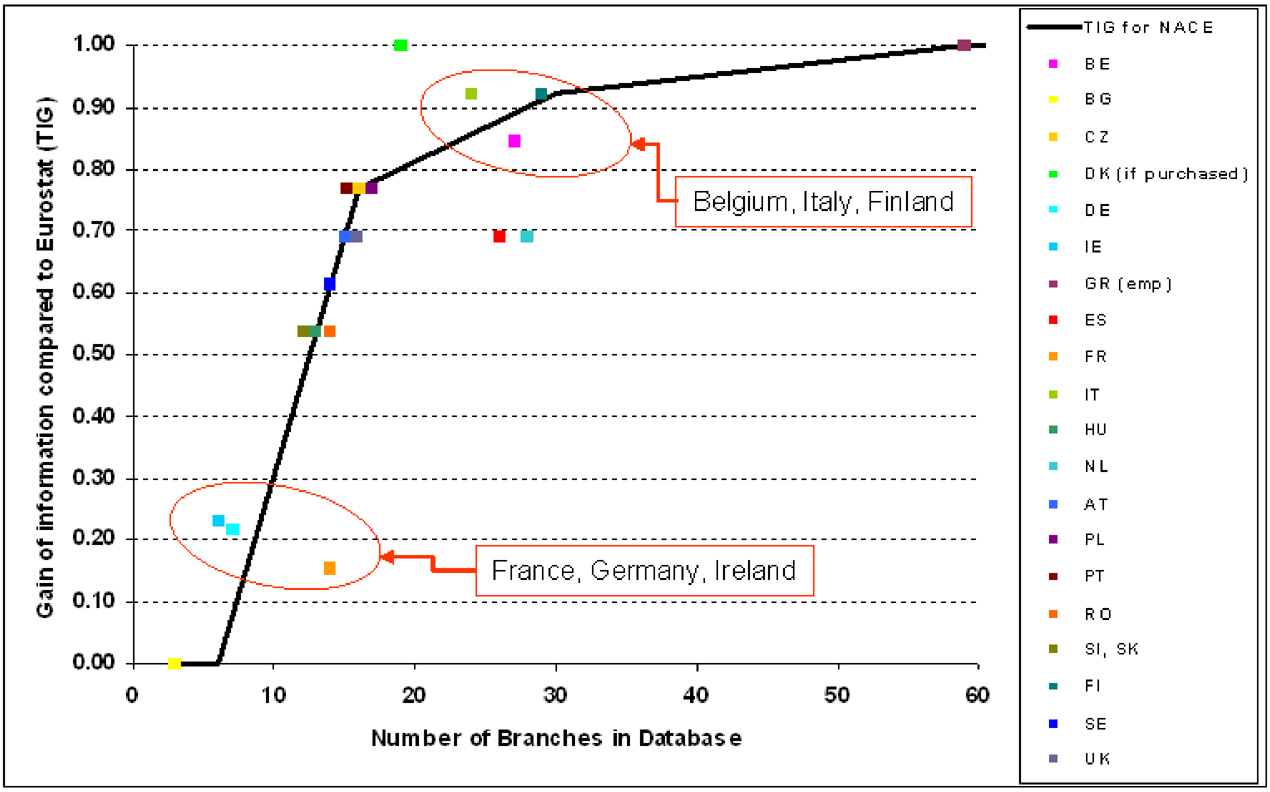
$$\delta\left(\sum_{AR}G_{i,b19,AR}^{19\_AR}\right)=\begin{cases}1 & if \sum_{AR}G_{i,b19,AR}^{19\_AR}>0\\ 0 & if \sum_{AR}G_{i,b19,AR}^{19\_AR}=0\end{cases} \tag{2}$$

The aggregator matrix G^{19_AR} is constructed by assigning ones to branches that can be mapped in a many-to-one way to the 19 branches, and zeroes otherwise. The indicator can range between 0 (no informational gain compared to ESTAT data) and 1, which indicates a full coverage of branches for the respective indicator. The TIG measures only the potential gain of information counting only the usable branch classifications in the national datasets.
Screening the national statistical departments’ supply resulted in a rather mixed picture. For free datasets, the informational gain as measured by the TIG-Indicator never overcomes 0.92 (Figure 1). The smallest informational gain is observed for Bulgaria (no gain) and France, Germany, and Ireland, for which the range of gains is below 0.3 points. One reason is the aggregate representation of agriculture, forestry, and fishery in the regional branch accounts

intermediation; real estate, renting and business activities", L_TO_P; "Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies"

as for 12 Member States. As a disaggregation of these branches is crucial, the availability of other indicators that could permit a split has been evaluated.
The second purpose of screening the national statistical organisations was to retrieve the availability of regional IOTs. The effort has been modestly satisfactory as survey-based regional IOTs had been obtained for 12 Comunidades Autonomas (NUTS2) of Spain and the NUTS1 regions Scotland and Baden-Württemberg. The IOTs obtained for Finland and those potentially available for Austria, Italy, and Poland are mainly based on non-survey methods.

Figure 1 Branches in regional accounts for b1g, and informational gain compared to ESTAT data – only Member States with more than one NUTS2 region



Notes:

- The black line indicates the gain of information for the different NACE (A6, A16, A30, A60)
- Scale of informational gain indicator: 0: Same information as obtained from ESTAT; 1: Full coverage of all IOTNUTS2 branches

Source: Own presentation

The focus of the survey of the EU Member States' statistical departments is on the availability of regional branch account indicators, like gross value-added and employment. Availability of additional datasets has been scrutinized, i.e datasets on population and migration as indicators for the mobility of labour within and across the EU Member States. Population figures are available for all NUTS2 regions from ESTAT and national statistical organisations, while migration figures are not as easily obtainable. The NSO provide more information than ESTAT, with the exception of Denmark, Greece, France, Malta, and Portugal.
After the compilation of the inventory, we construct a full set of a-priory SAMs based on national SAM coefficients and regional employment and value-added data. The compilation

and consolidation of these indicators for all 271 NUTS2 regions based on the obtained data is discussed in the following section.

5. Performance of Non-survey Methods

The regional branch accounts ($RAMA^{ESTAT}$) for compensation of employees, employment and gross value added, despite their rough aggregation, are selected as the benchmark to derive the regional a-priori SAMs. The main advantage of these indicators is the completeness in terms of spatial coverage. Once selected the benchmark, three datasets ($RAMA^{ESTAT}$, national branch account tables $NAMA^{ESTAT}$ and branch accounts from the national statistical organisations $RAMA^{NSO}$) have to be combined. The NSOs do not always provide full information for the B19 branch classification but rather for an A16 scheme where Agriculture and Forestry and Manufacturing are combined. Accordingly, the information for three branch classifications (A6, A16, and B19) needs to be combined. This is done by aggregator matrices (G) between the branch classifications and the usage of shares of B19 and A16 entries in the respective A6 (and A16) entries. The following example illustrates the procedure for the simplified situation without the A16 classification.
Using this aggregator matrix, the national shares of B19 entries in A6 are calculated for the national datasets (NASH) for all member states (MS), where i represents the indicators from the NAMA datasets, t the years 2000 to 2005. G^{19_6} represents the aggregator matrix between A6 NACE and B19 target branches:

$$NASH_{b19,i,t}^{ESTAT,MS} = \frac{NAMA_{b19,i,t}^{MS}}{\sum_{A6}\left(G_{b19,A6}^{19\_6} \cdot \sum_{b19}\left[G_{b19,A6}^{19\_6} \cdot NAMA_{b19,i,t}^{MS} \right]\right)} \tag{3}$$

Similarly, regional shares in A6 are calculated (RASH):

$$RASH_{b19,i,t}^{NSO,R} = \frac{RAMA_{b19,i,t}^{NSO,R}}{\sum_{A6}\left(G_{b19,A6}^{19\_6} \cdot \sum_{b19}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{b19,i,t}^{NSO,R} \right]\right)} \tag{4}$$

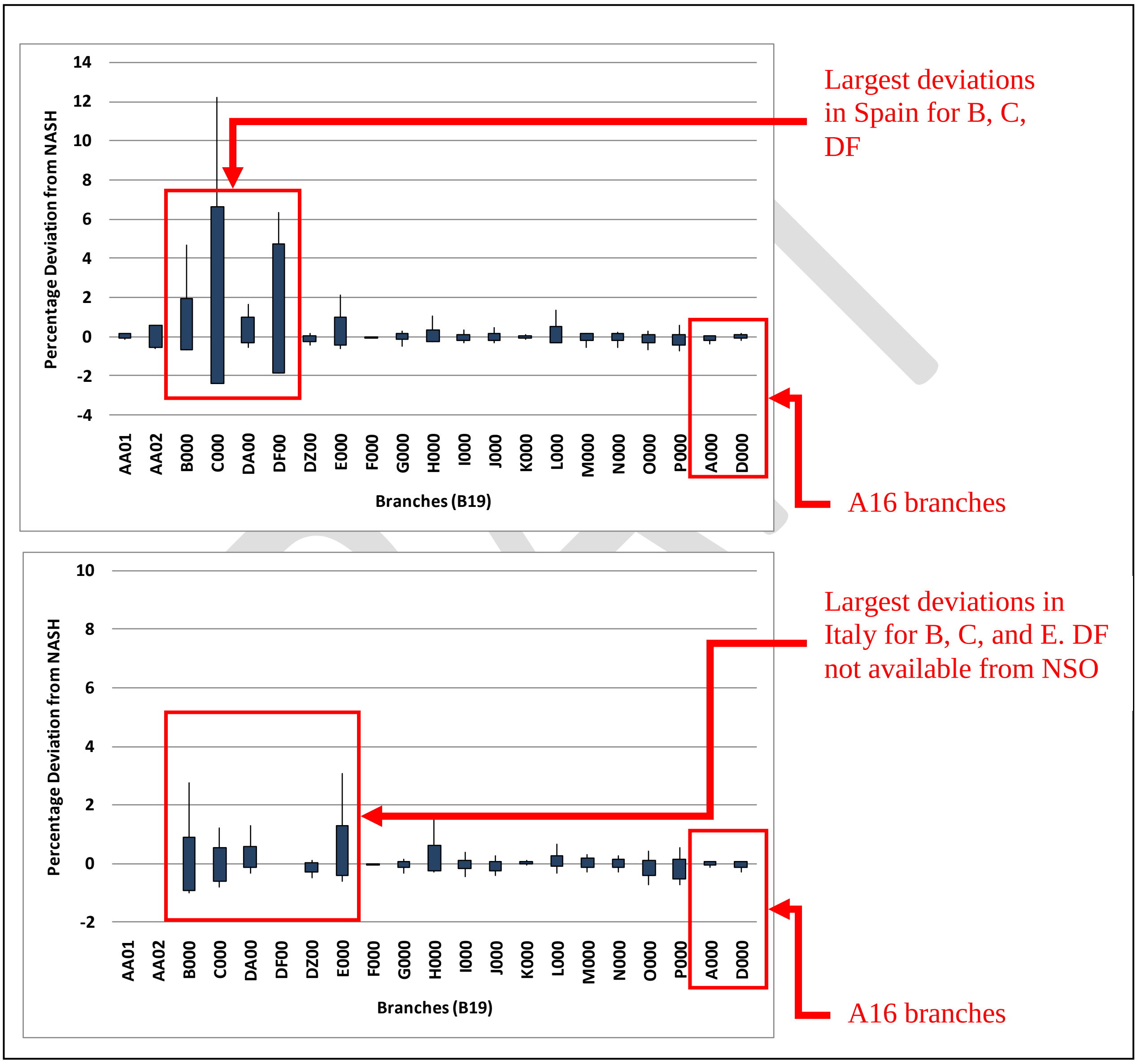
Combination into a core-account dataset Y (regional indicators, A: intermediate demand, V: value-added, L: labour indicators) is then achieved by an expansion of the $RAMA^{ESTAT}$ datasets and multiplication with the shares, depending on the availability of datasets from NSO:

$$Y_{b19,i,t}^{MS,R} = \begin{cases} \sum_{A6}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{A6,i,t}^{ESTAT,R} \right] \cdot RASH_{b19,i,t}^{NSO,R} & \forall RAMA_{b19,i,t}^{NSO,R} \neq 0 \\ \sum_{A6}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{A6,i,t}^{ESTAT,R} \right] \cdot NASH_{b19,i,t}^{ESTAT} & \forall RAMA_{b19,i,t}^{NSO,R} = 0 \end{cases} \tag{5}$$

The deviation between NASH and RASH figures for Italy and Spain, for which regional data are available from NSOs, are then compared (Figure 2). The regional branch composition within the A6 aggregates does not deviate from the national shares, with few exceptions. Represented in the case of Spain by sub-sectors of A2B (particularly B: Fisheries) and C2E (particularly C: Mining and Quarrying, as well as DF: Fuel industry). For Italy, the largest deviation can be observed for fisheries, mining and quarrying and electricity production. These observations are understandable as primary production, like Fisheries and Mining, and

energy sectors, like Fuel industries and Electricity, are usually concentrated in some regions (close to the sea, natural resources), while they may not exist in other regions.
The shares of D (Manufacturing) and A (Agriculture and Forestry) do not deviate to the same extent from the national shares and permit to derive at least for the A16 a reliable set of core accounts. For the share of D in C2E, D (at least in Spain) has a high value, such that biased estimates for C and E have not a huge impact on the regional economies.

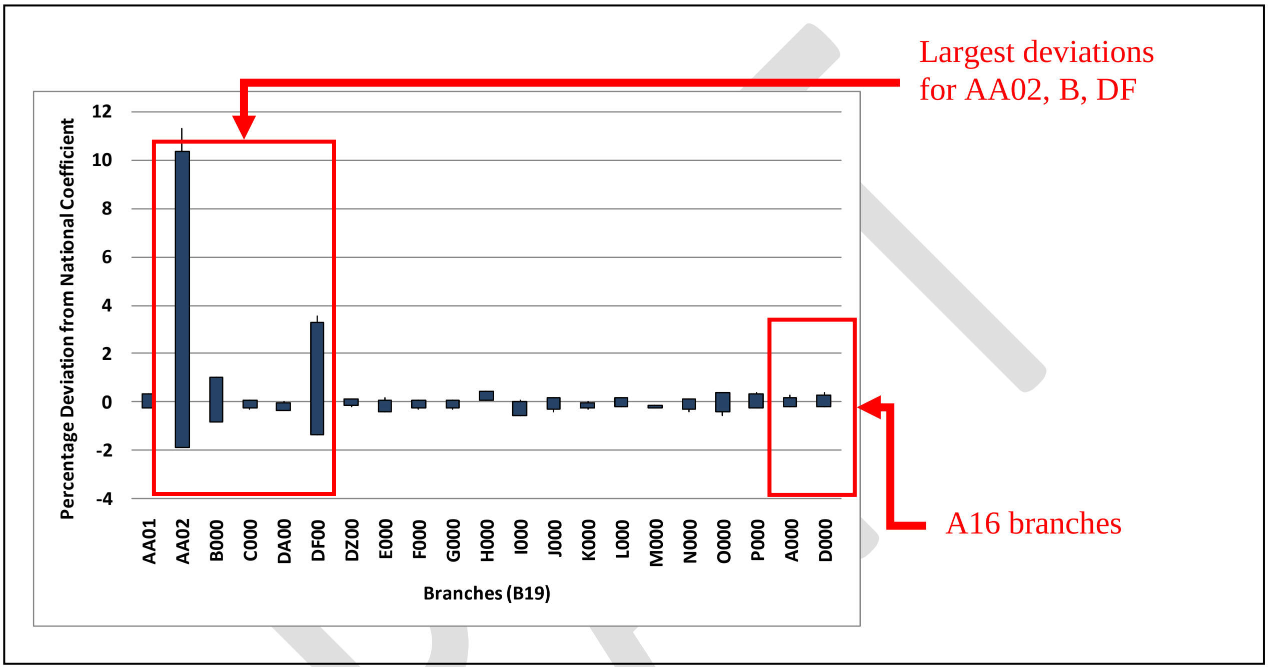
Figure 2 Deviation between National and Regional Shares in Spain and Italy



The proposed derivation of regional core accounts for employment and gross value-added based on ESTAT regional branch accounts and national shares of B19 in A6 branch aggregates produce acceptable results when considering the regional economies (dominated by primary production or manufacturing). Nevertheless, detailed information on the share of fisheries and mining and quarrying industries significantly improve the reliability of the picture of the regional economies.
To complete the regional core accounts and derive regional IOTs, additional indicators, as gross output by branch, total intermediate demand by branch, or taxes on activities paid by branch are derived. The default procedure is to use national coefficients for the completion of

the core accounts. We assume that per unit of gross value-added generated by economic branch, a similar share of intermediate input would be used and a similar share of gross output would be produced as on a national scale. Figure 3 compares the outcomes of these computations with the recorded figures from the Spainish NSO. The largest deviations for intermediate inputs occur for forestry, fisheries and most notably fuel industries. This result can also be observed for indicators like gross output and consumption of fixed capital.

Figure 3 Deviation between National and Regional Coefficients for Intermediate Inputs (p2)



These observations clarify that forestry, fisheries, and the fuel industry have a different regional cost structure compared to the national averages. One reason could be that policy measures like subsidies for some inputs are implemented on a regional scale. As the compared figures refer to output and intermediate demand at basic prices, and not to physical units, the distortions caused by regional policies may be severe. In the absence of additional information, we assume regionally different tax and subsidy rates and account for this in the final balancing steps.

Comparing the results obtained from these steps with observed branch account data from NSO, we observed deviations in the regional composition of the A2B and C2E branch aggregates from the national averages, particularly for forestry, fisheries, mining and quarrying, and fuel industries. This observation is not surprising as these branches tend to be concentrated and are branches which constitute regional economic heterogeneity. For those branches additional information has been collected. For industrial sectors, Structural Business Statistics proved to be a valuable, even if limited, source. Data on fisheries and forestry could not be obtained in a comparable manner, but the availability of CAPRI data for agriculture and the fact that the share of the branch aggregate “agriculture and forestry” within the A2B aggregate is often uniformly distributed across regions facilitated the derivation of consistent branch account data for forestry as residual. Based on this, figures for fisheries could also be derived residually.

For the industrial sectors, the majority of NSO provided regional, disaggregated branch account data, although with varying levels of detail. The informational gain from NSOs for 15 Member States with more than one NUTS2 region permit an improvement of at least 25%

compared to the use of ESTAT branch accounts and national shares and that the split into the targeted 5 sub-sectors (C,da,df,dz,E) can be reliable performed.

6. Derivation of I-O Tables and Balancing Procedure

After the derivation of regional core accounts, the regional SAMs are compiled. The literature established a widely accepted sequence of procedural steps, entitled the “Generation of Regional Input-Output Tables” procedure, or GRIT. GRIT draws mainly on the application of Location Quotients (LQs). LQs play a crucial role on several stages of the compilation sequence. The objective of LQs is to derive the regional intermediate demand-from-regional-origin by adjusting the national coefficients according to the weight a branch has in the regional economy, usually measured by employment shares of the branch in the region. We assume that a high weight of a branch in a region causes the development of other regional branches that may provide the needed intermediate inputs. This is plausible for intermediates with high transportation costs, thus making regional branches competitive compared to suppliers from other regions. (Flegg et al., 1995, or Bonfiglio and Chelli, 2008). The two Flegg Location Quotients (Flegg LQ and Augmented Flegg LQ) not only depend on sectoral national and regional employment data, but also on the choice of the parameter δ ($0 \leq \delta < 1$), which introduces an "element of flexibility" (Flegg et al. 1995). The choice of δ depends on empirical considerations (Flegg at al 1995), and statistical properties are investigated by Bonfiglio and Chelli (2008), finding higher values for δ to yield better results based on a Monte Carlo Analysis.

Location Quotients performance is tested for the Spanish IOTs. We calculate the root mean squared error (RMSE) for each LQ and divide by the RMSE of the default setting LQ=1. An improvement over the simple usage of national unadjusted coefficients is achieved (all values smaller than 1). However, the best performance is observed applying an AFLQ with a δ of 0.4, in line with the findings of Bonfiglio and Chelli (2008), who observed the best performance of AFLQ with a δ between 0.3 and 0.5.

The database includes sub-matrices for intermediates from domestic and imported origin. The resulting need to estimate international and inter-regional trade flows due to the lack of recorded data is a widely recognized challenge for multi-regional modelling. Gravity models are usually employed to solve this problem. Gravity models help to derive total intermediate demands from domestic or imported origin, but not to derive the full sub-matrices. To estimate the full set of required sub-matrices, we derive intermediate domestic and imported demand using total national intermediate demand coefficients. Spanish data shows that although the goodness of fit for the sub-matrices are not entirely satisfying because of systematic deviation between derived and recorded values, the correlation between derived and recorded value is still high. Having calculated the demand for regionally produced intermediates based on Location Quotients and total intermediate use and imported intermediate, the sub-matrix of domestic imported intermediate use is computed residually.

Having generated a priori entries for the production accounts of the target IOT, the value-added is distributed across the receiving institutions: government and households. Private households are the primary recipients of wages and salaries and gross operating surplus, while the local government receives indirect taxes and employers' social contributions. The aggregate private income is then used to determine direct taxes and transfers based on national tax and transfer rates obtained from the NASA dataset. On the demand side, due to the lack of information on regional consumption expenditures, we derive private regional consumption based on national consumption rates. Finally, the ESTAT dataset “reg_e2gfcf” provides information for 175 NUTS2 regions, sometimes overlapping with data from

RAMA^{NSO} datasets on aggregate regional investment. Where no information is available, we resort to the usage of investment shares in national income.
The database developed following the previous steps does not automatically fulfil the requirement that regional expenditures equal regional revenues, nor that the regional IOTs add up to the national ones. Furthermore, the tables have to be consistent with control totals. Only in few circumstances we observe the true values of the targeted tables, but rather the distorted ones. Thus, a robust estimator must be used to consolidate the IOTs. The observed values should be as close as possible to the real, fulfilling the usual constraints of equality between row and columns, the add-up conditions and the consistency with control totals. Identified these boundary conditions, a statistical criterion that allows estimating the new values, as close as possible to the observed ones, should be identified.
Round (2003) provides an overview on SAM and IOT estimation and balancing approaches, including Generalised Cross Entropy (GCE, Golan et al 1994, Robinson et al 2001), Stone-Byron (Stone 1977, Byron 1978), and RAS (Bacharach 1970). Many variants and combinations of the basic methods are applied, some building on column-coefficients (Breisinger et al. 2007, Robinson et al. 2001), some on the table entries (Mueller 2006, Nakamura 1998). Control totals on the tables sub-matrices are imposed rigorously (Robinson et al 2001), or can be associated with an error term (Breisinger et al. 2007). The procedure selected to balance depends on the available data and on how the researcher expresses his trust in the information at hand. For the IOTNUTS2, the procedure has to build on the entries rather than column coefficients to ensure compliance with control-totals and prior information at regional level.
A Generalized Cross Entropy application, as proposed by Golan et al. (1994) and Robinson et al. (2001) requires for each estimate matching priors and weights, and with them two supports and may cause computational difficulties for large-scale datasets. Additionally, the implicit posterior density depends on the interaction between the choice of supports, the a priori probabilities and the entropy criterion. Both problems are addressed by Heckelei et al. (2008) and Witzke and Britz (2005) by motivating a Highest Posterior Density (HPD) estimator which refrains from discrete support points but still allows to express confidence by using informative priors on the variance of each estimate. In HPD approaches, the objective function is minimized subject to constraints, an approach similar to the Stone-Byron one, although motivated by Heckelei et al. (2008) from a different perspective. This setting creates a substantially smaller computational burden, not only because of fewer constraints and variables, but also because of the linearity of the first derivative of objective function. Round (2003) suggests that despite many studies on updating SAMs seem to prefer GCE methods, an extended Stone-Byron method which include additional constraints, should be favoured over other methods. Its advantage is to allow incorporating judgement on the relative reliability of data sources. Therefore this method is closer to the spirit of the problem. Because of these reasons, the application of a HPD rather than a GCE procedure for the IOTNUTS2 database is preferred.
We therefore split the balancing procedure into several steps, following the steps of the prior datasets construction. Starting point is the balancing of national core accounts for all target sectors and sector-aggregates, then of national SAMs and finally the regional ones.

7. Conclusions

This paper outlined several findings from an ongoing project on the compilation of regional SAMs for the European Union at NUTS2 level

References

Bacharach, M. (1970), Biproportional Matrices and Input-Output Change. Cambridge University Press, London.

Bonfiglio, A. and F. Chelli (2008): Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input-Output Tables through a Monte Carlo Simulation. In: Economic Systems Research, 20(3): 243-258

Breisinger, C., J. Thurlow and D. Magnus (2007): A 2005 Social Accounting Matrix (SAM) for Ghana. Chana statistical Services; International Food Policy Research Institute.
Britz, W. and H.P. Witzke (eds.) (2008) CAPRI Model Documentation2008: Version 2. URL: http://www.capri-model.org/docs/capri_documentation.pdf, University of Bonn.

Britz, W. and H.P. Witzke (eds.) (2008) CAPRI Model Documentation2008: Version 2. URL: http://www.capri-model.org/docs/capri_documentation.pdf, University of Bonn.

Byron, R. P. (1978): The estimation of Large Social Account Matrices. In: Journal of the Royal Statistical Society, Series A, 141: 359–367.

Casini-Benvenuti, S. and R. Panicciá (2003): A Multi-Regional Input-Output Model for Italy, Istituto Regionale Programmazione Economica Toscana (IRPET), Interventi, note e rassegne n. 22/03 (http://www.irpet.it/storage/pubblicazioneallegato/34_Interventin.22.pdf).

ESTAT (2008): ESTAT Manual of Supply, Use and Input-Output Tables. URL: http://epp.ESTAT.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-07-013

Flegg A. T., C. D. Webber, and M. V. Elliott (1995): On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables. In: Regional Studies 29: 547- 561

Fritz, O., R. Kurzmann, G. Zakarias and G. Streicher (2003): Constructing Regional Input-Output Tables for Austria. In: Austrian Economic Quarterly, 1/2003.

Golan, A., G. Judge, and S. Robinson (1994): Recovering Information from Incomplete or Partial Multisectoral Economic Data. In: Review of Economics and Statistics, 76(3): 541–49.

Heckelei, T., R. Mittelhammer, and T. Jansson (2008): A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models. Agricultural and Resource Economics, Discussion Paper 2008:2, Bonn, Germany.

Jansson, T., M.H. Kuiper, M. Adenäuer (2009) Linking CAPRI and GTAP, SEAMLESS Report No.39, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 100 pp. ISBN no. 978-90-8585-127-1.

Mattas, K., F. Arfini, P. Midmore, M.Schmitz and Surry Y. (2011): The impact of the CAP on regional employment: a multi-modelling cross-country approach. In: Disaggregated Impacts of CAP Reforms: Proceedings of an OECD Workshop, OECD Publishing.

Mueller, M. (2006): A General Equilibrium Approach to Modeling Water and Land Use Reforms in Uzbekistan. Dissertation University of Bonn.

Müller, M., I. Pérez Domínguez, and S. H. Gay (2009): Construction of Social Accounting Matrices for EU27 with a Disaggregated Agricultural Sector (AgroSAM). Luxembourg: Office for Official Publications of the European Communities. EUR – Scientific and Technical Research series ISBN 978-92-79-13625-2

Nakamura, Y. (1998), Investment and Saving in Russian Macroeconomy: Compilation and Analyses of an Aggregated SAM for Russia, 1995. CERT Working Paper No. DP98/09.

Nowicki, P., V. Goba, A. Knierim, H. van Meijl, M. Banse, B. Delbaere, J. Helming, P. Hunke, K. Jansson, T. Jansson, L. Jones-Walters, V. Mikos, C. Sattler, N. Schlaefke, I. Terluin and D. Verhoog (2009) Scenar2020-II – Update of Analysis of Prospects in the Scenar2020 Study. European Commission, Directorate-General Agriculture and Rural Development, Brussels.

Official Statistics of Finland (OSF), (2011): Regional input-output tables [e-publication]. Helsinki: Statistics Finland, http://www.stat.fi/til/apt/index_en.html.

Robinson, S., A. Cattaneo, and M. El-Said (2001): Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods. In: Economic Systems Research, 13(1): 47-64.

Round, J. (2003): Constructing SAMs for Development Policy Analysis: Lessons Learned and Challenges Ahead. In: *Economic System Research*, 15(2).

Rueda-Cantuche, J. M., J. Beutel, F. Neuwahl, I. Mongelli, and A. Loeschel (2009): A Symmetric Input-Output Table for EU27: Latest Progress. In: Economic Systems Research 21(1):59-79

Stone, R (1977): Foreword to G. Pyatt & A. R. Roe (with R.M. Lindley, J.I. Round and others), Social Accounting for Development Planning (Cambridge, Cambridge University Press), pp. xvi-xxxi.

Törmä, H. and K. Zawalinska (2007): Technical description of the CGE RegFin/RegPol models. University of Helsinki, RURALIA Institute. URL: http://www.helsinki.fi/ruralia/research/manuals.htm

Törmä, H., K. Zawalinska, M. Blanco-Fonseca, E. Ferrari and T. Jansson (2010): Regional CGE model layout with a focus on integration with the partial equilibrium models and modelling of RD measures, CAPRI-RD Deliverable 3.2.1, (http://www.ilr1.uni-bonn.de/agpo/rsrch/capri-rd/docs/d3.2.1.pdf)

Witzke, H.-P. and W. Britz (2005): Consolidating trade flows and market balances globally using a Highest Posteriori Density estimator. Presented at the 8th Annual Conference on Global Economic Analysis, Lübeck, Germany

Appendix 1 – Structure of the Social Accounting Matrices

S ≡

	b	d	1	w	1
bw	**Aw**	**Cw**	**Iw**	**Ew**	
b	**A**	**C**	**I**	**E**	**X**
f	**F**	**T**			**0**
f+1	**Ts**				**0**
1	**X**	**0**	**0**	**0**	

Where:

b: Index for economic branches
d: Index for regional institutions like private households and governments
w: Index for trade partners or origins of the items in the SAM (regional, domestic, foreign)
f: Index for primary factors of production (labour, land, capital)
S: Regional Social Accounting Matrix
Aw: Intermediate demand, distinguished by origin
A: Total intermediate demand ($A_{b,b'} = \sum_{w} Aw_{b,w,b'}$)
Cw: Final demand, distinguished by origin
C: Total final demand ($C_{b,d} = \sum_{w} Cw_{b,w,d}$)
Iw: Investment demand, distinguished by origin
I: Total investment demand ($I_b = \sum_{w} Iw_{b,w}$)
Ew: Exports, distinguished by origin and destination (e.g. exports of imported goods to other regions of the same country)
E: Total exports by destination
F: Payments to fixed factors ($E_{b,w} = \sum_{w'} Ew_{b,w',w}$)
Ts: Taxes on production and factors
T: Transactions between institutions (distribution of regional income)
X: Total regional supply (has to equal total regional demand):

$$X_{b'} = \sum_{b} A_{b,b'} + \sum_{f} \left[F_{f,b'} + Ts_{f,b'} \right] + Ts_{*,b'}$$

$$= \sum_{b} A_{b',b} + \sum_{d} C_{b',d} + I_{b'} + \sum_{w} E_{b',w}$$

Appendix 2 – Structure of the Satellite Accounts

$$\mathbf{Y} \equiv \left(\begin{array}{c|ccccc|} & b & d & 1 & w & 1 \\ \hline w & \multicolumn{4}{c}{\mathbf{M}} & \\ 1 & \mathbf{At} & \mathbf{Ct} & \mathbf{It} & \mathbf{Et} & \\ 1 & \mathbf{V} & & & & \mathbf{Vt} \\ 1 & \mathbf{L} & & \mathbf{P} & \mathbf{G} & \\ \hline \end{array}\right)$$

Where:

Y: Regional Satellite Accounts
M: Total imports by branch or institution and by origin
At: Total intermediate demand in the region ($At_{b'} = \sum_b A_{b,b'}$)
Ct: Total final demand by institution (households, governments) ($Ct_d = \sum_b C_{b,d}$)
It: Total regional investment ($It = \sum_b I_b$)
Et: Total regional exports by destination
V: Gross value-added by branches ($V_b = \sum_f \left[F_{f,b} + Ts_{f,b} \right] + Ts_{*,b}$)
Vt: Total regional gross value-added (Gross Regional Product) ($Vt = \sum_b V_b$)
L: Employed persons by branch
P: Regional population
G: Net-migration to or from abroad or other regions of the same country